University of Dundee

Gathering Point-Aided Viral Marketing in Decentralized Mobile Social Networks

Fan, Bo; Leng, Supeng; Yang, Kun; He, Jianhua

Published in:
IEEE Systems Journal

DOI:
10.1109/JSYST.2016.2520206

Publication date:
2018

Document Version
Peer reviewed version

Gathering Point-Aided Viral Marketing in Decentralized Mobile Social Networks

Bo Fan, *Student Member, IEEE*, Supeng Leng, *Member, IEEE*, Kun Yang, *Senior Member, IEEE*, and Jianhua He, *Senior Member, IEEE*

***Abstract*—Viral marketing is a technique that spreads advertisement information through social networks. Recently, viral marketing through online social networks has achieved huge commercial success. However, there are still very little research reported on viral marketing in decentralized mobile social networks (MSNs). Comparing with online viral marketing, viral marketing in decentralized MSNs faces many challenges, such as unreliable information diffusion and limited network knowledge. To address these problems, we propose the *gathering point-aided mobile viral marketing* (*GP-MVM*) scheme, which contains two major components, i.e., *seed selection* and *information diffusion*. *Seed selection* is responsible to select a set of seed nodes from which information diffusion begins. Based on a new metric called integrated contact strength (ICS), we propose two distributed seed selection schemes, i.e., *ratio seeding* and *threshold seeding*, while, for information diffusion, we propose the *GP-aided diffusion* algorithm, which utilizes user GPs to promote information propagation. Continuous-time Markov chain-based analytical model shows that GP-MVM has a good scalability. Simulations indicate that GP-MVM outperforms two state-of-the-art information diffusion methods designed for MSNs, in terms of both diffusion proportion and diffusion speed.**

***Index Terms*—Gathering point (GP), information diffusion, integrated contact strength (ICS), mobile social networks (MSNs), viral marketing.**

I. Introduction

MOBILE social networks (MSNs) [1] consist of mobile nodes with social characters, such as interest and social relationship. Basically, MSN can be classified into two categories, i.e., centralized MSN [2] and decentralized MSN [3]. Centralized MSN, also called web-based MSN, is the extension of online social networks (OSNs) with users accessing to the OSN services through mobile devices, while a decentralized MSN contains a group of users that forward information through opportunistic encounters. Information is forwarded through short-range communication, such as WiFi-direct, without connecting to a centralized sever. Therefore, decentralized MSN is also regarded as a type of delay tolerant network [4]. In this paper, we focus on decentralized MSN.

Viral marketing is a technique that spreads advertisement information through social networks. Taking the advantage of "word-of-mouth" spreading, viral marketing can quickly diffuse information from a single node to the whole network, like the propagation of viruses. Based on the social relationship of users in OSN, online viral marketing (OVM) can diffuse information among friends and further to the friends of friends, as shown in Fig. 1(a). Formulated as an influence maximization problem, OVM has been well investigated [5]–[7]. However, viral marketing through decentralized MSN, which we call mobile viral marketing (MVM), is rarely studied. Different from OVM, MVM relies on the geographical neighborship of nodes [Fig. 1(b)], since nodes in MSN send information through short-range communication. However, as the neighborship of nodes in MSN often changes, information diffusion is unreliable. In addition, since globe knowledge of the network is usually unavailable in MSN, solving the influence maximization problem is not a feasible strategy for MVM. Both these challenge the design of MVM approaches.

Considering these challenges, this paper proposes a distributed scheme, i.e., the *gathering point-aided MVM* (*GP-MVM*) scheme, to realize efficient MVM in MSN. Similar to plenty of existing viral marketing schemes [2], [5]–[7], GP-MVM consists of two components, i.e., *seed selection* and *information diffusion*.

Seed selection aims at selecting a set of seed nodes, from which information can be quickly diffused to the whole network. In recent studies, seed selection is operated at the goal of maximizing information influence, with a globe knowledge of the network. However, this is often infeasible for MVM as globe knowledge is difficult to be obtained in a decentralized MSN. Hence, we propose two distributed schemes for seed selection, i.e., *ratio seeding* and *threshold seeding*, which only employ the local knowledge of each node.

Information diffusion defines how to diffuse information from the seed nodes to the network. Two basic diffusion models are often adopted in viral marketing, i.e., 1) the linear threshold model [6] and 2) independent cascade model [8]. However, both of them are only used to compute the influence maximization problem. Besides, they are also not applicable for a decentralized MSN scenario. Fortunately, information diffusion in MSN has been well investigated, which can provide insight to the design of MVM schemes. Many recent researches [3], [9] study

Manuscript received August 31, 2015; revised December 17, 2015; accepted January 13, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61374189, in part by the Fundamental Research Funds for the Central Universities, China, under Grant ZYGX2013J009, in part by EU FP7 Project CLIMBER under Grant GA-2012-318939, in part by U.K. EPSRC Project DANCER under Grant EP/K002643/1, and in part by the National High-Tech R&D Program under Grant 2015AA015701. *(Corresponding author: Supeng Leng.)*

B. Fan and S. Leng are with the School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: spleng@uestc.edu.cn).

K. Yang is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: kunyang@essex.ac.uk).

J. He is with the School of Engineering and Applied Science, Aston University, Birmingham B4 7ET, U.K. (e-mail: j.he7@aston.ac.uk).

Digital Object Identifier 10.1109/JSYST.2016.2520206

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

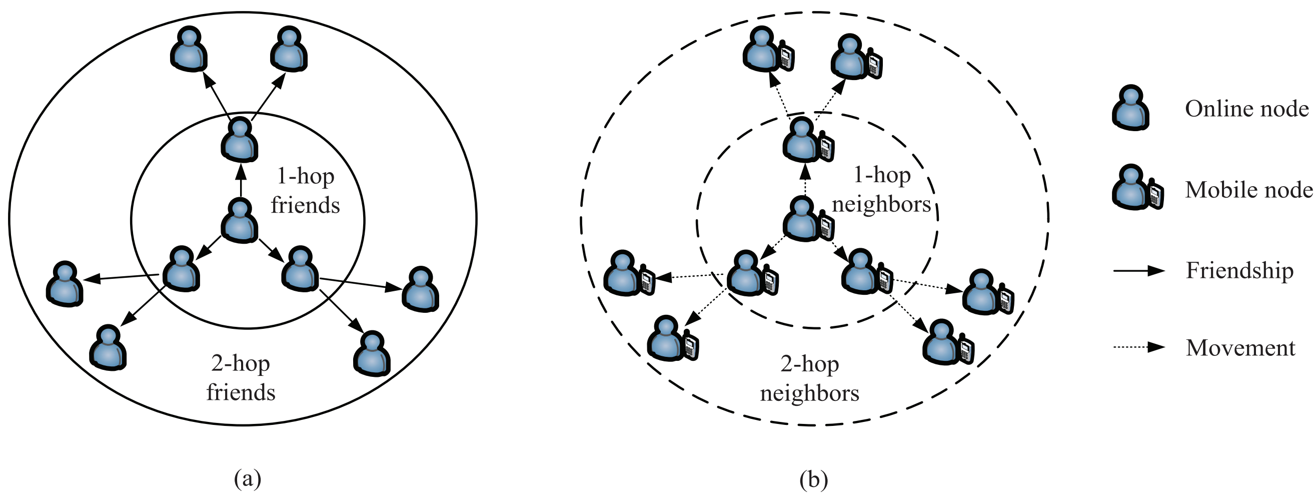


Fig. 1. (a) OVM and (b) MVM diffuse information via social relationship and geographical neighborship, respectively.

the utilization of human GPs in information diffusion. GPs are popular places often visited by a large number of nodes, such as supermarket. A GP can act as a relay node by equipping a wireless device to receive, store, and forward information. GPs are naturally good relay nodes because they can bridge many nodes for information diffusion. Hence, we employ GPs as relay nodes and propose the *GP-aided diffusion* algorithm. Different from the existing methods, which only exploit the social strength among nodes [10]–[12], we further consider the social strength between nodes and GPs. Consequently, forwarding information toward GPs can be more efficient, which ultimately increases the information diffusion speed. The main contributions of this paper can be summarized as follows.

1) We propose a new metric called integrated contact strength (ICS) for seed selection and seed refreshing, which exploits not only the social strength among nodes but also the one between nodes and GPs. Based on ICS, the selected seed nodes can quickly diffuse information to other nodes or store it at the GPs for larger forwarding chances.
2) We propose the GP-MVM scheme which has two components: seed selection and information diffusion. Two distributed seed selection schemes, i.e., ratio seeding and threshold seeding, are developed. As for information diffusion, the GP-aided diffusion algorithm is devised.
3) A Markov chain-based analytical model is proposed for analyzing the property of the GP-MVM scheme. The contributions of the GPs and the seed nodes in information diffusion are compared. Besides, the scalability of GP-MVM is studied.

This paper is organized as follows. Section II provides a review of the related work. Section III presents the preliminary of the work, followed by the detail of the GP-MVM scheme in Section IV. Sections V and VI present the theoretical analysis and the simulations of GP-MVM, respectively. Finally, Section VII concludes this paper.

II. Related Work

A. Viral Marketing

Viral marketing aims at spreading the influence of product through social networks. Viral marketing was first studied by Domingos and Richardson [13] as an influence maximization problem. But they only provided a probabilistic solution of the problem. Kempe *et al.* [6] formulated the problem as an optimization problem and proposed a greedy algorithm to solve it. However, it was revealed by Chen *et al.* [14] that this algorithm is slow and unscalable. Instead, they proposed a heuristic algorithm that gains a much faster spreading speed by restricting computation on the local influence regions of nodes. Wang *et al.* [7] proposed a community-based greedy algorithm to solve the influence maximization problem quickly. Work [5] and [15] considered the cost of the method and, respectively, proposed a cost-effective scheme. In work [2], instead of maximizing influence, the authors tried to minimize the information diffusion time. Nevertheless, all these methods focus on an OSN or centralized MSN scenario, where the globe knowledge of the network is available for calculating the influence maximization problem.

B. Information Diffusion in MSN

Information diffusion is one of the most challenging issues in decentralized MSN. *Epidemic routing* [16] is a simple information diffusion scheme in which information is diffused through flooding. It achieves a good performance but has a large cost of network resource. To solve this problem, *Spray&Wait* [17] restricts the number of nodes forwarding information. However, without utilizing any network knowledge, it cannot determine which node is better in diffusing information and select qualified relay nodes. As an improved version of *Spray&Wait*, *Homing Spread* [3] takes the advantage of human GPs called "home" for information diffusion. However, similar to Spray&Wait, Homing Spread also fails to employ any network knowledge. *SocialCast* [10] and *SimBet* [11] are two knowledge-based methods that select the next hop of information according to the social attributes of nodes. SocialCast exploits the one-hop neighbors and the varying pattern of the neighbor list, while SimBet employs the betweenness centrality and the similarity with the destination node. These approaches only utilize the social strength among nodes. The one between nodes and specific places, such as GPs, is often ignored.

III. Preliminary

A. MVM Features and Network Model

MVM diffuses information through geographical neighborship rather than social relationship. Hence, information is mainly diffused to the nodes that are geographically closed to the information source. This makes MVM much more useful than OVM in some situations. For example, a shopping center can diffuse its advertisement to the people who live nearby through this way, instead of using OSNs. However, due to the properties of decentralized MSN, MVM still has some disadvantages comparing with OVM.

First, as globe knowledge of the network is usually unavailable in decentralized MSN, information diffusion is based on a local knowledge of each nodes. Moreover, nodes in MSN are limited with storage and power, which makes information diffusion, a nonignorable cost for the nodes. Consequently, the cost of information diffusion must be controlled, by restricting the lifecycle of information, the number of nodes diffusing information and so forth. Finally, due to the movement of nodes, the probabilities of nodes for forwarding information are varying with time. Therefore, the nodes diffusing information should be selected and refreshed adaptively according to the social and network status of nodes.

In this paper, we consider all the shortcomings mentioned above and attempt to develop an efficient MVM scheme that can overcome these shortcomings. We consider a decentralized MSN scenario that contains 1 information source, n nodes, and m GPs. The information source tries to diffuse information to all the nodes to realize MVM. Instead of forwarding information to the nodes directly, the information source diffuses information to a set of seed nodes, which store the information and further forward it to other nodes. Information is forwarded through short-range communication, without resorting to base stations. Each GP is assumed to support a throwbox [18], a storage device with wireless interface, to act as a relay node. When a seed node reaches a GP, it can store information at the throwbox. Then, the throwbox can forward the information to other later coming nodes.

B. Integrated Contact Strength

Before introducing the detail of the GP-MVM schemes, we first develop a metric for seed selection and seed refreshing, by employing the social strength among nodes and GPs. Social strength means the strength in terms of social relationship, contact, etc. In this paper, we adopt contact strength as the metric. For two nodes, contact strength indicates the frequency they encounter each other. While for a node and a GP, contact strength means the frequency that the node visits the GP.

A good seed node should have the capability of forwarding information to both the nodes and the GPs. Hence, it should have a large contact strength with both the nodes and the GPs. Therefore, we define the metric as the *ICS* of a node with the nodes and GPs. The ICS of an arbitrary node i can be calculated via three steps. First, the contact strength between node i and a single node/GP is calculated. Then, the contact strength between node i and all the other nodes or all the GPs is computed. Finally, the ICS of node i with all the other nodes and all the GPs is derived. Contact history of node i that contains the history that it encounters other nodes and visits the GPs is explored as *a priori* knowledge.

1) Contact Strength With a Single Node/GP: From the contact history, multiple contact characters such as duration, frequency, and intercontact time can be obtained. According to work [19], among the contact characters, the average intercontact time is more comprehensive for evaluating contact strength, because it can reflect the properties of both duration and frequency. Consequently, we define the contact strength C_{ij}^N between node i and node j and the contact strength C_{ik}^G between node i and GP k by

$$C_{ij}^N = \frac{1}{D_{ij}^N}$$
$$C_{ik}^G = \frac{1}{D_{ik}^G} \tag{1}$$

where D_{ij}^N is the average intercontact time between node i and node j, while D_{ik}^G indicates the average intercontact time between node i and GP k.

2) Contact Strength With All the Nodes/GPs: In order to calculate the contact strength between node i and all the other nodes or all the GPs, the probability density function (pdf) of the intercontact time should be known. We assume that the intercontact time between two nodes and the one between a node and a GP, respectively, follow exponential distributions:

$$f_{ij}^N(t) = C_{ij}^N e^{-C_{ij}^N t}, \quad t > 0$$
$$f_{ik}^G(t) = C_{ik}^G e^{-C_{ik}^G t}, \quad t > 0 \tag{2}$$

with expectations D_{ij}^N and D_{ik}^G. Although it is indicated by some researches [20] that the power-law distribution may approximate the intercontact time of human better, the exponential distribution is still a good approximation and is widely employed by many researches [3], [9], [21], [22]. Moreover, the unique properties of the exponential distribution make it much easier for theoretical analysis.

To calculate the contact strength of node i with all the other nodes, we can regard all the other nodes as a group and compute the intercontact time between node i and the group (namely, the interval for node i to encounter any one in the group), whose pdf can be calculated as

$$f_i^N(t) = d[1 - P(T_1 > t, \ldots, T_j > t, \ldots, T_n > t)]/dt$$
$$= d\left(1 - \prod_{j=1, j\neq i}^{n} e^{-C_{ij}^N t}\right) \Big/ dt$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$$= d\left(1 - e^{-\sum_{j=1,j\neq i}^{n} C_{ij}^{N} t}\right)\Big/ dt$$
$$= \sum_{j=1,j\neq i}^{n} C_{ij}^{N} e^{-\sum_{j=1,j\neq i}^{n} C_{ij}^{N} t} \tag{3}$$

where T_j is a state variable indicating the intercontact time between node i and node j. The expectation of it is $1/\sum_{j=1,j\neq i}^{n} C_{ij}^{N}$. Hence, according to the definition of contact strength in (1), the contact strength between node i and all the other nodes can be derived by

$$C_i^N = \sum_{j=1,j\neq i}^{n} C_{ij}^N. \tag{4}$$

Similarly, we can derive the contact strength between node i and all the GPs as

$$C_i^G = \sum_{k=1}^{m} C_{ik}^G. \tag{5}$$

3) Derivation of ICS: To calculate the ICS of node i, we can regard all the other nodes and all the GPs as a group. The pdf of the intercontact time between node i and the group can be derived with the similar idea as (3)

$$f_i(t) = \left(\sum_{j=1,j\neq i}^{n} C_{ij}^N + \sum_{k=1}^{m} C_{ik}^G\right) e^{-\left(\sum_{j=1,j\neq i}^{n} C_{ij}^N + \sum_{k=1}^{m} C_{ik}^G\right)t}. \tag{6}$$

Hence, the ICS of node i can be derived by

$$C_i = \sum_{j=1,j\neq i}^{n} C_{ij}^N + \sum_{k=1}^{m} C_{ik}^G = C_i^N + C_i^G \tag{7}$$

which is an unweighted sum of C_i^N and C_i^G. However, this may be not reasonable in some situations. For example, visiting a popular GP that is frequently visited by other nodes is much more efficient than encountering a single node for information diffusion. So, the two contact strengths should have different weights. In this case, we modify the ICS function as

$$C_i = \omega C_i^N + (1-\omega) C_i^G, \quad \omega \in (0,1) \tag{8}$$

which allows for adjusting the importance of C_i^N and C_i^G. According to (8), the ICS is defined as a tradeoff between C_i^N and C_i^G, rather than the sum of them.

IV. Gathering Point-Aided Mobile Viral Marketing

With the aid of the metric ICS, we propose the *GP-MVM* scheme. Similar to other viral marketing approaches, GP-MVM is composed of two components, i.e., seed selection and information diffusion. For seed selection, we develop two distributed seed selection schemes, i.e., *ratio seeding* and *threshold seeding*. As for information diffusion, we devise the *GP-aided diffusion* algorithm.

A. Distributed Seeding Schemes

The first step of GP-MVM is selecting a particular number c of seed nodes. Seed selection starts from the information source and proceeds in a distributed way.

Binary spray [17] is a simple strategy for seed selection. According to binary spray, in order to select c seed nodes, the information source first duplicates the information into c copies and acts as the first seed. When a seed encounters another node that has no information copy, it sends half of its copies to the node and makes it a new seed. This process goes on until the c copies are diffused to exactly c seed nodes. Binary spray is fast because it simply selects the first c encountered nodes as seed nodes. However, as no metric is used for selection, the selected seed nodes may be unqualified for diffusing information because they cannot frequently encounter other nodes or visit the GPs. This can increase the whole information diffusion time. Moreover, in binary spray, a seed may store or send multiple same copies, leading to unnecessary wastage of device resource. Instead, it only needs to store/send one copy and use a label to indicate the number of copies (NC) it stores/sends.

With metric ICS, we modify binary spray and propose two new schemes, i.e., ratio seeding and threshold seeding. Each node maintains a label called number of copies (NC) to indicate the NC it has. Only seed nodes have $NC > 0$. Both schemes start from the information source who has $NC = c$.

Definition 1 (Ratio Seeding): For an arbitrary seed h with $NC_h > 1$, if encountering another node h' who satisfies $NC_{h'} = 0$ and $\lfloor NC_h \cdot \frac{C_{h'}}{C_h + C_{h'}} \rfloor > 0$, h sends a copy to h' and makes

$$NC_{h'} = \left\lfloor NC_h \cdot \frac{C_{h'}}{C_h + C_{h'}} \right\rfloor$$
$$NC_h = \left\lceil NC_h \cdot \frac{C_h}{C_h + C_{h'}} \right\rceil. \tag{9}$$

NC_h, $NC_{h'}$, C_h, and $C_{h'}$ are the NC and ICS of h and h', respectively. Ratio spray ensures that information copies are mainly taken by nodes with large ICS.

Definition 2 (Threshold Seeding): For an arbitrary seed h with $NC_h > 1$, if encountering another node h' who satisfies $NC_{h'} = 0$ and $C_{h'} > C_{threshold}$, h sends an information copy to h' and makes

$$NC_{h'} = \lfloor NC_h/2 \rfloor$$
$$NC_h = \lceil NC_h/2 \rceil. \tag{10}$$

$C_{threshold}$ is a threshold guaranteeing that the ICS of the selected seed is large enough. Both ratio seeding and threshold seeding terminate when c nodes have $NC = 1$.

B. GP-Aided Diffusion Algorithm

After seed selection, information can be diffused through the GP-aided diffusion algorithm, which is composed of three phases, i.e., *seed refreshing*, *information storing,* and *information diffusion.*

In *seed refreshing* phase, seed nodes are refreshed (i.e., nodes with larger ICS replace the seed nodes with smaller ICS). Seed

Algorithm 1. Gathering Point-aided Diffusion

```
Seed Refreshing
  for each seed h with NC_h = 1 do
      if encountering node h' that has NC_{h'} = 0 and
      C_{h'} > C_h
          sends an information copy to h' and makes
          NC_h = 0, NC_{h'} = 1;
Information Storing
  for each seed h do
      if reaching GP h' then
          stores an information copy at h';
Information Diffusion
  for each GP h do
      if node h' arrives then
          sends an information copy to h';
  for each seed h with NC(h) = 1 do
      if encountering node h' then
          sends an information copy to h';
```

refreshing is seldom considered in existing OVM schemes [5], [6], as the number of seed nodes is not restricted. However, considering the limited storage and power of nodes in MSN, the number of seed nodes should be restricted to control the cost of information diffusion. Moreover, in order to overcome the time-varying properties of nodes, the seed nodes should be refreshed adaptively according to the network and social status of nodes. As shown in Algorithm 1, when a seed h with $NC_h = 1$ encounters another node h', it compares their ICS values. If h' has a larger ICS value than h, h will send its only information copy to h' and let it be the new seed.

In *information storing* phase, information is stored at the GPs by the seed nodes for further diffusion. According to recent studies [3], [9], utilizing GPs as relay nodes emerges as a promising way to improve the efficiency of information diffusion in MSN. Some GP-assisted information diffusion approaches have been proposed [3], [9]. In these approaches, seed nodes are selected randomly or selected according to the social strength among nodes. The social strength between nodes and GPs (i.e., the strength that the nodes visit the GPs) is not considered. Hence, the selected seed nodes may not be able to store information at the GPs quickly. In GP-MVM, seed nodes are selected and refreshed according to the social strength between nodes and GPs (i.e., ICS). Therefore, the selected seed nodes are more capable for information storing.

In *information diffusion* phase, the seed nodes and the GPs diffuse information to the nodes that have not received it. A node can receive information from a seed when they encounter each other or from a GP when it reaches the GP. Note that, only the seed nodes with $NC = 1$ can conduct information diffusion. For seed nodes with $NC > 1$, seed selection will be executed.

Basically, seed selection starts before information diffusion. But they can be executed in parallel. For example, when a seed encounters a node with a larger ICS, if the seed has $NC > 1$, seed selection is preformed. Otherwise, seed refreshing is executed. Moreover, the three phases in the diffusion algorithm are also executed in parallel. The definition of the two components and the three phases is only for the convenience of illustration.

V. Theoretical Analysis

In this section, we study the properties of the GP-MVM scheme using an analytical model based on a continuous-time Markov chain [23].

A. Basic Settings

The network size is denoted by $\mathcal{B}$, which is linearly related with the numbers of nodes n and GPs m in the network. For example, if the network size $\mathcal{B}$ doubles, the numbers of nodes and GPs will both double. The total contact strength of each node is a constant. To simplify the analysis, the intercontact times between any two nodes and between a node and a GP follow exponential distributions with parameters λ and Λ, respectively,

$$\begin{aligned} f_N(t) &= \lambda e^{-\lambda t}, \quad t > 0 \\ f_G(t) &= \Lambda e^{-\Lambda t}, \quad t > 0. \end{aligned} \tag{11}$$

λ and Λ denote the contact strengths between any two nodes and between a node and a GP. λ and Λ are set as constants in some studies, such as [3], since the network size $\mathcal{B}$ is set as a constant. However, when $\mathcal{B}$ is a variable, in order to keep the total contact strength of each node as a constant, we set $\lambda = \frac{\lambda'}{n}$ and $\Lambda = \frac{\Lambda'}{m}$, respectively, where λ' is a constant representing the total contact strength of a node with other nodes, while Λ' indicates the total contact strength of a node with the GPs.

B. Continuous-Time Markov Chain

The information diffusion process is modeled as a continuous-time Markov chain (Fig. 2), whose state space $\mathcal{S}$ is defined as

$$\mathcal{S} = \{s_{ij} = \langle i, j \rangle \mid 0 \le i \le m, 0 \le j \le n\} \tag{12}$$

where i and j indicate the numbers of GPs and nodes that have received the information, respectively. The transitions among the states in $\mathcal{S}$ are denoted by

$$\mathcal{T} = \{T(s, s') \mid s \in \mathcal{S}, s' \in \mathcal{S}, s \neq s'\} \tag{13}$$

where $T(s, s')$ indicates one-step transition from state s to state s'. Note that, both s and s_{ij} are used to indicate a certain state. s is for more general indication, while s_{ij} is used when the indexes i and j are discussed. Instead of discrete probability [24], we employ the pdf $f_{s,s'}(t)$ for each state transition $T(s, s')$, where t is the waiting time before the transition takes place. $f_{s,s'}(t)$ satisfies

$$f_{s,s'}(t) \begin{cases} > 0, & \text{if } s \text{ and } s' \text{ are adjacent,} \\ = 0, & \text{otherwise} \end{cases} \tag{14}$$

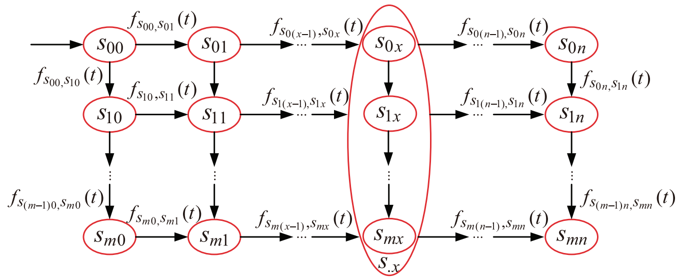


Fig. 2. State transition diagram of the Markov chain.

where being adjacent means if s is s_{ij}, then s' is either $s_{(i+1)j}$ or $s_{i(j+1)}$.

The state transition begins from the initial state s_{00} where information diffusion starts and terminates at state s_{mn} where all the m GPs and the n nodes have received the information. The state transition takes place when and only when a GP or a node receives the information. The pdf of each transition can be calculated with Theorem 1.

Theorem 1: For an arbitrary state s_{ij}, if state transition $T(s_{ij}, s_{(i+1)j})$ exists, the pdf $f_{s_{ij},s_{(i+1)j}}(t)$ for transition $T(s_{ij}, s_{(i+1)j})$ equals

$$c(m-i)\Lambda e^{-(c(m-i)\Lambda+c(n-j)\lambda+i(n-j)\Lambda)t}. \tag{15}$$

Similarly, if state transition $T(s_{ij}, s_{i(j+1)})$ exists, the pdf $f_{s_{ij},s_{i(j+1)}}(t)$ for transition $T(s_{ij}, s_{i(j+1)})$ equals

$$(c(n-j)\lambda+i(n-j)\Lambda)e^{-(c(m-i)\Lambda+c(n-j)\lambda+i(n-j)\Lambda)t}. \tag{16}$$

Proof: See Appendix A.

C. Information Diffusing Time

Based on the Markov chain, the information diffusion time can be obtained. Instead of calculating the time for all the nodes to receive the information, we compute the expected time for a specific proportion of nodes to receive the information. In other words, with the total number of nodes specified, we compute the expected time for a specific number x of nodes to receive the information. That is the expected transition time from the initial state s_{00} to any of the states $\{s_{ix} \mid 0 \le i \le m\}$. We define, by $g_{s,s'}(t)$, the pdf that state transits from s and first reach s'. Clearly, $g_{s,s'}(t) = f_{s,s'}(t)$ if s and s' are adjacent. The computation of information diffusion time consists of three steps. First, for each state s_{ij}, $g_{s_{00},s_{ij}}(t)$ (abbreviated as $g_{ij}(t)$) is computed. Then we regard the states $\{s_{ix} \mid 0 \le i \le m\}$ as a compound state $s_{.x}$ and calculate $g_{s_{00},s.x}(t)$ (abbreviated as $g_{.x}(t)$). Finally, the expected information diffusion time is calculated. To avoid confusion, several pdfs used in this section is sorted in Table 1.

1) Derivation of $g_{ij}(t)$: The state transition from s_{00} to s_{ij} can take place through different paths with specific probabilities. It is difficult to consider all the paths directly. Therefore, we adopt a recursive method.

TABLE I
NOTATIONS

$f_{s,s'}(t)$	PDF for one-step transition from s to s'
$g_{s,s'}(t)$	PDF for state transition from s to s' (first reach)
$g_{ij}(t)$	PDF for state transition from s_{00} to s_{ij}
$g_{.x}(t)$	PDF for state transition from s_{00} to any of $s_{.x}$

Algorithm 2. Compute $g_{ij}(t), \forall i = 1, \ldots, m, j = 1, \ldots, n$

1: Given the state transition diagram as Fig. 2;
2: Compute $f_{s_{ij},s_{(i+1)j}}(t)$ and $f_{s_{ij},s_{i(j+1)}}(t)$ for each state s_{ij} based on Eq. (15) and Eq. (16);
3: $i = 0$;
  For $j = 2$ to n
    $g_{ij}(t) = \int_0^t g_{i(j-1)}(\tau) f_{s_{i(j-1)},s_{ij}}(t-\tau)d\tau$;
4: $j = 0$;
  For $i = 2$ to m
    $g_{ij}(t) = \int_0^t g_{(i-1)j}(\tau) f_{s_{(i-1)j},s_{ij}}(t-\tau)d\tau$;
5: For $i = 1$ to m
    For $j = 1$ to n
      $g_{ij}(t) = \int_0^t g_{(i-1)j}(\tau) f_{s_{(i-1)j},s_{ij}}(t-\tau)d\tau$
        $+ \int_0^t g_{i(j-1)}(\tau) f_{s_{i(j-1)},s_{ij}}(t-\tau)d\tau$;

Theorem 2: Consider two arbitrary states s and s' and the previous states of s', $SP_{s'} = \{s'' \mid f_{s'',s'}(t) > 0, s'' \in \mathcal{S}\}$. $g_{s,s'}(t)$ can be calculated by

$$g_{s,s'}(t) = \sum_{s'' \in SP_{s'}} \int_0^t g_{s,s''}(\tau) f_{s'',s'}(t-\tau)d\tau. \tag{17}$$

Proof: See Appendix D.

Theorem 2 indicates that if the pdfs of time for state transitions from state s to the previous states of s' are known, the pdf of the time for transition from s to s' can also be derived. Consequently, we propose Algorithm 2 to compute $g_{ij}(t)$ for each state s_{ij}. The computational complexity of this algorithm is $O(|\mathcal{S}|^2)$. Since $|\mathcal{S}| = (m+1)(n+1)$, it can be also denoted as $O(m^2n^2)$.

2) Derivation of $g_{.x}(t)$: We treat the states $\{s_{ix} \mid 0 \le i \le m\}$ as a compound state $s_{.x}$ which represents that x nodes have received the information. Consider the previous states of $s_{.x}$, $SP_{s_{.x}} = \{s_{i(x-1)} \mid 0 \le i \le m\}$, similar to Theorem 2, $g_{.x}(t)$ can be derived by

$$g_{.x}(t) = \sum_{0 \le i \le m} \int_0^t g_{i(x-1)}(\tau) f_{s_{i(x-1)},s_{ix}}(t-\tau)d\tau. \tag{18}$$

As $g_{i(x-1)}(t)$ for $\forall s_{i(x-1)} \in SP_{s_{.x}}$ is calculated by Algorithm 2, $g_{.x}(t)$ can be derived with (18).

3) Expected Information Diffusion Time: As $g_{.x}(t)$ is derived, the expected time for x nodes to receive the information can be calculated by

$$D_x = \int_0^\infty t g_{.x}(t)dt. \tag{19}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

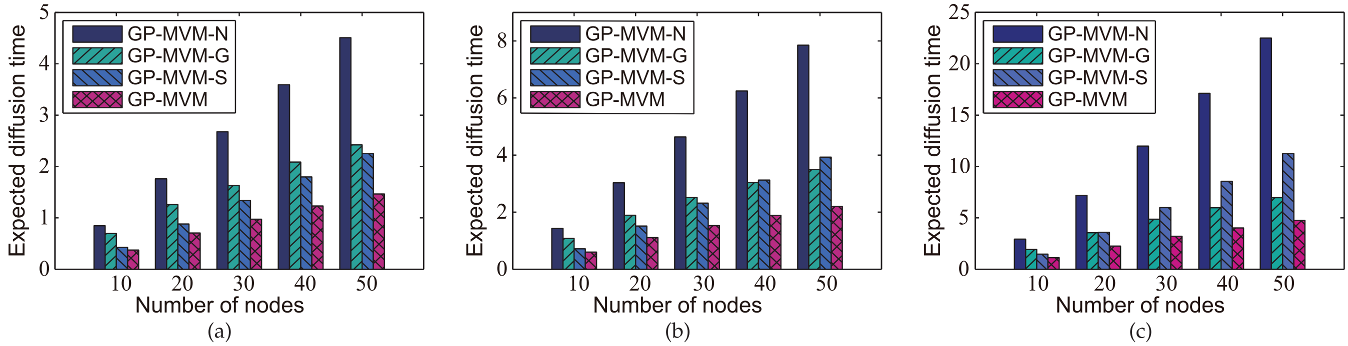


Fig. 3. Comparison of the four scenarios. (a) Diffusion proportion = 60%. (b) Diffusion proportion = 80%. (c) Diffusion proportion = 100%.

D. *Numerical Results and Analysis*

Now we study the properties of GP-MVM with the expected information diffusion time.

1) Contributions of Seed Nodes and GPs: Since both seed nodes and GPs help diffusing information, we first study the contributions that they each make. With $\lambda' = 10$ and $\Lambda' = 1$, we compare four schemes, including 1) *GP-MVM-N*, in which neither seed nodes nor GPs are leveraged, and information is diffused by the information source itself; 2) *GP-MVM-S*, where seed nodes diffuse information but do not store it at the GPs; 3) *GP-MVM-G*, in which only the information source diffuses and stores information; and 4) *GP-MVM*, where both seed nodes and GPs are leveraged. We study the performance of the four schemes under different network sizes. Since the numbers of nodes n and GPs m are linearly related with each other, we define $n = 5$ m. The maximum number of seed nodes c is set to be 2. The expected diffusion times for 60%, 80%, and 100% of nodes to receive the information are calculated.

As shown in Fig. 3, the expected diffusion times of the four schemes are calculated with the number of nodes increasing. The GP-MVM-N scheme unsurprisingly has the largest diffusion time among the four schemes. GP-MVM-S obtains a smaller diffusion time that is half of GP-MVM-N, since the number of seed nodes is 2 and the delivery probability is doubled. With the assist of GPs, GP-MVM-G also gets a smaller diffusion time. Note that the expected diffusion time of GP-MVM-G is larger than GP-MVM-S when the network is small. However, when the network grows to a certain size, it becomes close to or even smaller than GP-MVM-S, indicating that the contribution of GPs increases with the network size. Finally, GP-MVM achieves the shortest expected diffusion time, since both seed nodes and GPs are leveraged. The comparison indicates that both seed nodes and GPs make an obvious contribution in improving the performance of information diffusion.

2) Scalability of GP-MVM: We next study the scalability of GP-MVM by analyzing how the cost of GP-MVM changes with the network size $\mathcal{B}$. The cost of GP-MVM is defined as the maximum number $C_{\max}$ of information copies stored in the network. According to GP-MVM, $C_{\max} = c + m$, where c is the maximum number of seed nodes, and m is the number of GPs in the network.

We increase the network size and compute the diffusion time for different proportions of nodes to receive information, with

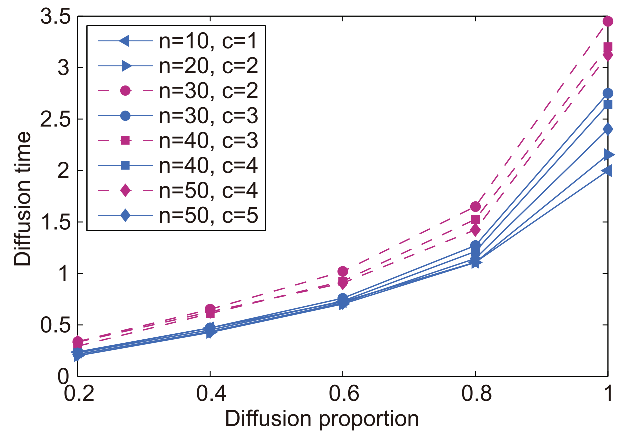


Fig. 4. Performance under different n and c.

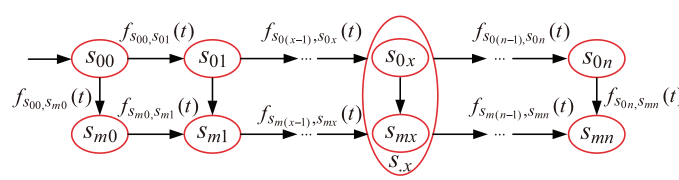


Fig. 5. State transition diagram for interconnected GPs.

different values of c. Due to the limitation of space, we only present eight cases as shown in Fig. 4. It can be found that, with the network size enlarged (n increased), the performance of GP-MVM degrades. For example, when $n = 30$ and $c = 2$, a larger diffusion time is suffered comparing with the case where $n = 20$ and $c = 2$. However, with the increase of c, GP-MVM can maintain a relatively steady performance even if the network size enlarges. Indeed, the five full lines are almost overlapped, indicating very similar diffusion times. Particularly, according to Fig. 4, when the number of nodes increases from 10 to 50, to keep a stable performance, c should increase from 1 to 5, indicating a linear relationship between c and n. In addition, since $m = \frac{1}{5}n$, $C_{\max} = c + m$ also has a linear relationship with n, indicating that the cost $C_{\max}$ is linearly related with the network size, i.e., $C_{\max} = O(\mathcal{B})$. Therefore, GP-MVM has a good scalability.

3) Disconnected GPs Versus Interconnected GPs: Finally, we study a special case where GPs are interconnected and share information. This is a common situation in real life. For example, in a college or an office building, the distributed WiFi

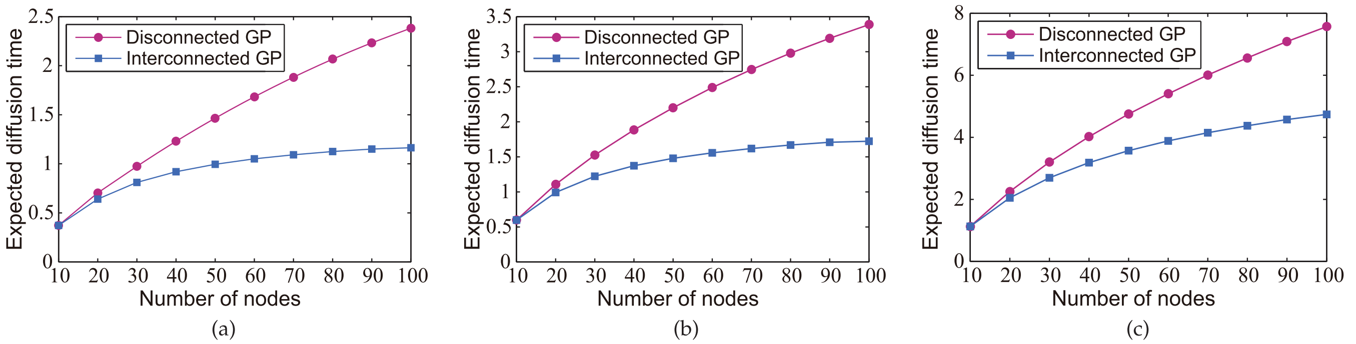


Fig. 6. Disconnected GP versus interconnected GP in terms of diffusion time. (a) Diffusion proportion = 60%. (b) Diffusion proportion = 80%. (c) Diffusion proportion = 100%.

access points (which can be used as throwboxes) are interconnected through a centralized switch. In this case, once a GP receives the information, all the other GPs will get it. The state transition diagram under this case is shown in Fig. 5, and the pdfs for the state transitions can be calculated according to Theorem 1.

The expected diffusion time of GP-MVM under this case is compared with the general case with disconnected GPs. As shown in Fig. 6, the expected diffusion time for 60%, 80%, and 100% of nodes to receive the information increases with the network size in both cases. However, when GPs are interconnected, the increasing rate becomes small, as the information storing phase is shortened. Moreover, when the network grows to a certain size, the increase of diffusion time becomes very slow in the scenario with interconnected GPs. It implies that the expected diffusion time in this case may converge to a certain value. However, since the closed formula of the expected diffusion time is not available, the theoretical proof of such a feature is also unavailable. Nevertheless, it is an interesting point for further study.

VI. Simulation Results

In this section, we study the performance of GP-MVM by comparing it with two state-of-the-art information diffusion approaches for MSN, i.e., Homing Spread [3] and SocialCast [10]. Homing Spread adopts the similar diffusion strategy as GP-MVM but fails to exploit the social strength of nodes for seed selection. SocialCast employs the social strength among nodes. However, the one between nodes and GPs is neglected.

A. Basic Settings

The simulations are conducted on OMNeT++ simulator. The Dartmouth trace [25], collected from a 5-year experiment conducted at the Dartmouth college, is utilized as the mobility model of nodes. In the experiment, numerous WiFi access points are allocated in the main buildings of the campus. Once a node connects/disconnects to/from an access point, it will be recorded in a log. We randomly select 60 nodes and their mobility traces from September 25 to October 24, 2003 and set them as the nodes in the simulations. One randomly selected node acts as the information source and periodically diffuses information to the 60 nodes. Each information is set a time-to-live (TTL). The buildings frequently visited by the 60 nodes are regarded as GPs. According to the selected traces, eight GPs are selected. The main simulation parameters are listed in Table II. The unique parameters of the two comparative methods are set according to the default setting in [3] and [10]. Nodes are set to be homogeneous, with the same interest, device type, power, and storage, to avoid the influence of such settings on the performance.

TABLE II
Simulation Parameters

Simulation parameters	Default value
Simulation time	1 month
Seed selection scheme	Ratio seeding
Information generation period	10 min
Number of nodes n	60
Number of GPs m	8
Number of seed nodes c	4
Weight ω in ICS function	0.3

B. Simulation Results and Discussion

1) Diffusion Proportion and Diffusion Speed: First, We study two basic performance metrics: diffusion proportion and diffusion speed. Diffusion proportion is the proportion of nodes that receive information within the TTL of information, which also indicates the reliability of the method, while diffusion speed indicates how fast the information is diffused.

The performance of GP-MVM is compared with SocialCast and Homing Spread with TTL increasing from 1 to 10 h. Different values of c are used. Fig. 7 shows the diffusion proportion of the three approaches, while Fig. 8 shows the distribution of diffusion time of all the received information. It can be found that GP-MVM achieves a larger diffusion proportion than other two approaches. Besides, in GP-MVM, most information is received within a relatively short time (4 h). This owes to the following reasons. First, GP-MVM selects and refreshes seed nodes according to the ICS of nodes, while SocialCast only employs the social strength among nodes and Homing Spread never exploits any knowledge. Second, GP-MVM takes the advantage of GPs. The analysis in Section V has indicated that GPs can make a significant contribution in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

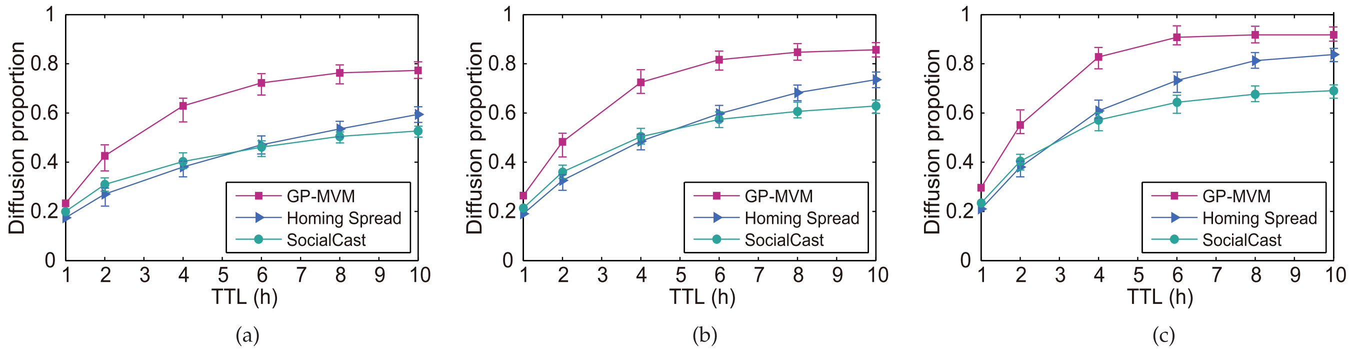


Fig. 7. Diffusion proportion comparison. (a) $c = 2$. (b) $c = 4$. (c) $c = 8$.

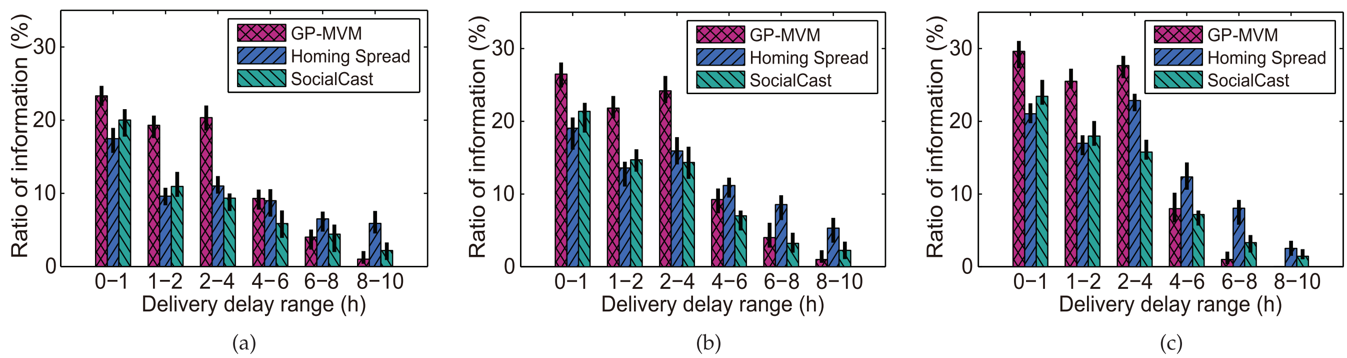


Fig. 8. Distribution of diffusion time of all received information. (a) $c = 2$. (b) $c = 4$. (c) $c = 8$.

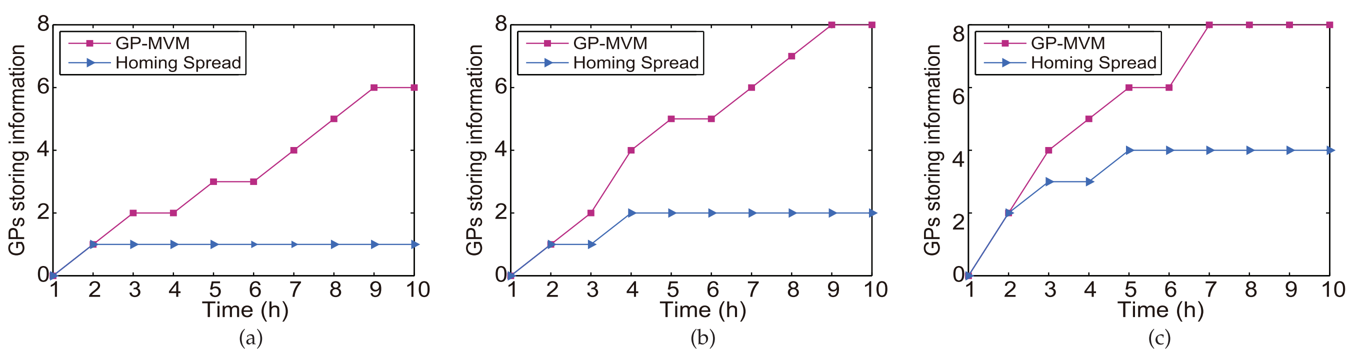


Fig. 9. Information storing comparison. (a) $c = 1$. (b) $c = 2$. (c) $c = 4$.

information diffusion. SocialCast does not make use of GPs, so it fails to take the advantage. Homing Spread employs GPs for information diffusion. However, no metric is used for seed selection or refreshing. Hence, information may be taken by nodes that seldom visit the GPs and information storing can be time consuming.

2) Information Storing: Information storing at GPs is an importance phase in both GP-MVM and Homing Spread. The efficiency of information storing directly determines the speed of the whole information diffusion process. Consequently, we compare the information storing phase of the methods. Since information storing is not applied in SocialCast, we only compare GP-MVM with Homing Spread.

A randomly chosen information is tracked to investigate the information storing process. Fig. 9 shows that GP-MVM is faster than Homing Spread in information storing. For example, when $c = 4$, it costs 5 h for Homing Spread to store the information at 4 GPs, while GP-MVM only uses 3 h. This is because that, seed nodes in GP-MVM are selected and refreshed based on the social strength between nodes and GPs. Hence, they are more likely to visit the GPs. In addition, in GP-MVM, the information is finally store by more GPs, because c in Homing Spread constrains the numbers of seed nodes and GPs that store information, while in GP-MVM, this number only constrains the number of seed nodes. Even if $c = 1$, GP-MVM can store information at all the GPs. This may bring GP-MVM a larger cost in storing and diffusing information comparing with Homing Spread. However, the analysis in Sections V-D2 and VI-B1 has proven that such a cost is tolerable and worthwhile for the gain in performance.

3) Weight in the ICS Function: The weight ω determines the relative importance of C_i^N and C_i^G in the ICS function. We study the influence of it on the performance of GP-MVM by increasing it from 0 to 1.

As shown in Fig. 10, with the increase of ω, both diffusion proportion and diffusion time decrease, indicating a

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

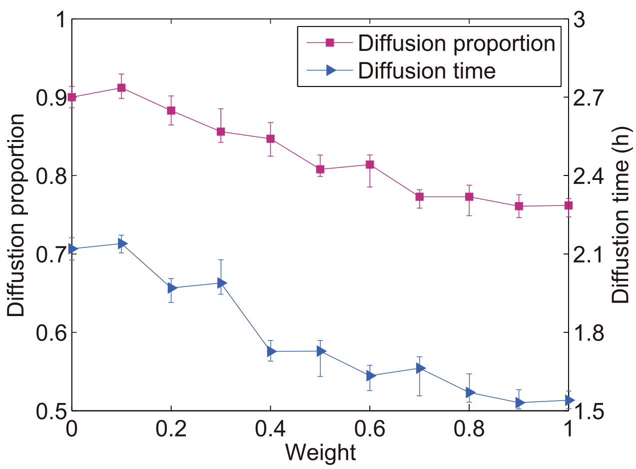


Fig. 10. Diffusion proportion and average diffusion time against weight ω.

degradation of diffusion proportion and an improvement of diffusion speed. when ω is relatively small (e.g., $\omega < 0.3$), GP-MVM performs well in diffusion proportion but poorly in diffusion speed. This is because the selected seed nodes are mainly those with a large contact strength with the GPs. They can spread the information to numerous GPs and obtain more delivery chances. However, since most diffusion chances are available after the information is stored, diffusion time becomes large. On the contrary, when the value of ω is relatively large (e.g., $\omega > 0.8$), GP-MVM experiences a short average diffusion time but achieves a small diffusion proportion as well. This is because that information is mainly taken by nodes with large contact strength with other nodes. Hence, it can be quickly diffused to the nodes. However, as the seed nodes may not visit the GPs frequently, some GPs are not able to get the information during the TTL of the information, and hence, many diffusion chances are lost.

The proper setting of ω depends on the demand of performance. If diffusion proportion is essential, ω should be set a small value. On the other hand, if the speed of information diffusion is required, ω should have a relatively large value. While, to balance the performance between diffusion proportion and speed, a medium value is suggested.

VII. Conclusion

This paper proposes the GP-MVM scheme to realize viral marketing in a decentralized MSN. GP-MVM consists of two components: seed selection and information diffusion. For seed selection, two distributed seed selection schemes, i.e., ratio seeding and threshold seeding, are proposed based on a new metric called ICS. As for information diffusion, the GP-aided diffusion algorithm is proposed, in which the human GP are utilized as relay nodes for information diffusion. Since the metric ICS employs the social strength between nodes and GPs, the selected seed nodes can quickly forward information to the GPs for storing and facilitate information diffusion. Continuous-time Markov chain-based analytical model indicates that GP-MVM has a good scalability. Simulations show that GP-MVM outperforms two state-of-the-art information

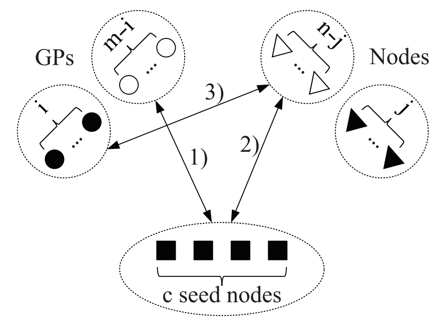


Fig. 11. State transitions from state s_{ij}.

diffusion methods designed for MSN, in terms of both diffusion proportion and speed. In our future work, we will further consider other issues, such as energy efficiency and security, for more efficient MVM.

Appendix A
Proof of Theorem 1

State s_{ij} indicates that i GPs and j nodes have received the information, and the remaining $m-i$ GPs and $n-j$ nodes have not yet (Fig. 11). The state transition occurs only when 1) any of the c seed visits any of the remaining $m-i$ GPs; 2) any of the c seed nodes encounters any of the remaining $n-j$ nodes; and 3) any of the remaining $n-j$ nodes visits any of the i GPs that have received the information. According to the property of the exponential distribution, the pdfs of the above three events, respectively, follow exponential distributions:

$$\begin{aligned} f_1(t) &= c(m-i)\Lambda e^{-c(m-i)\Lambda t} \\ f_2(t) &= c(n-j)\lambda e^{-c(n-j)\lambda t} \\ f_3(t) &= i(n-j)\Lambda e^{-i(n-j)\Lambda t}. \end{aligned} \tag{20}$$

The state transition $T(s_{ij}, s_{(i+1)j})$ occurs if and only if event 1 happens before events 2 and 3. According to the property of the exponential distribution, the pdf for transition $T(s_{ij}, s_{(i+1)j})$ can be given by

$$c(m-i)\Lambda e^{-(c(m-i)\Lambda + c(n-j)\lambda + i(n-j)\Lambda)t}. \tag{21}$$

Similarly, the transition $T(s_{ij}, s_{i(j+1)})$ occurs when event 2 or 3 happens before event 1. In this case, the pdf of transition $T(s_{ij}, s_{i(j+1)})$ is

$$(c(n-j)\lambda + i(n-j)\Lambda)e^{-(c(m-i)\Lambda + c(n-j)\lambda + i(n-j)\Lambda)t}. \tag{22}$$

Appendix B
Proof of Theorem 2

The transition from s to s' has to go through the previous states S_P of s'. Let us denote $|S_P| = z$ and $S_P = \{s_p^{(1)}, s_p^{(2)}, \ldots, s_p^{(z)}\}$. Then there exist z paths $Paths = \{path_i = s \rightarrow s_p^{(i)} \rightarrow s' \mid 1 \leq i \leq z\}$, through which the state can transit from state s to state s' with a specific probability for each path.

Let $g_{path_i}(t)$ denotes the pdf of the time for the transition from s to s' through $path_i$. Then, we have

$$g_{s,s'}(t) = \sum_{i=1}^{z} g_{path_i}(t). \tag{23}$$

Let the time of the transition from s to $s_p^{(i)}$ be τ. Then the time of the transition from $s_p^{(i)}$ to s' is $t - \tau$. τ can be any value between 0 and t. Hence, we have

$$g_{path_i}(t) = \int_0^t g_{s,s_p^{(i)}}(\tau) f_{s_p^{(i)},s'}(t-\tau) d\tau \tag{24}$$

$$g_{s,s'}(t) = \sum_{i=1}^{z} \int_0^t g_{s,s_p^{(i)}}(t) f_{s_p^{(i)},s'}(t-\tau) d\tau \tag{25}$$

which can be also expressed as

$$g_{s,s'}(t) = \sum_{s'' \in SP_{s'}} \int_0^t g_{s,s''}(\tau) f_{s'',s'}(t-\tau) d\tau. \tag{26}$$

References

[1] N. Kayastha, D. Niyato, P. Wang, and E. Hossain, "Applications, architectures, and protocol design issues for mobile social networks: A survey," *Proc. IEEE*, vol. 99, no. 12, pp. 2130–2158, Dec. 2011.

[2] Z. Lu, Y. Wen, and G. Cao, "Information diffusion in mobile social networks: The speed perspective," in *Proc. IEEE INFOCOM*, 2014, pp. 1932–1940.

[3] M. Xiao, J. Wu, and L. Huang, "Home-based zero-knowledge multi-copy routing in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1238–1250, May 2015.

[4] Z. Li, W. Li, S. Guo, X. Liu, S. Lu, and D. Chen, "On gateway selection in disruption tolerant networks," *Ad Hoc Sensor Wireless Netw.*, vol. 20, pp. 223–243, 2014.

[5] T. N. Dinh, H. Zhang, D. T. Nguyen, and M. T. Thai, "Cost-effective viral marketing for time-critical campaigns in large-scale social networks," *IEEE/ACM Trans. Netw.*, vol. 22, no. 6, pp. 2001–2011, Dec. 2014.

[6] D. Kempe, J. M. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. ACM SIGKDD*, 2003, pp. 137–146.

[7] Y. Wang, G. Cong, G. Song, and K. Xie, "Community-based greedy algorithm for mining top-K influential nodes in mobile social networks," in *Proc. ACM SIGKDD*, 2010, pp. 1039–1048.

[8] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Lett.*, vol. 12, no. 3, pp. 211–223, 2001.

[9] B. Fan, S. Leng, K. Yang, and Q. Liu, "GPS: A method for data sharing in mobile social networks," in *Proc. IFIP Netw.*, 2014, pp. 1–9.

[10] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, 748–760, Jun. 2008.

[11] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant MANET," in *Proc. ACM MobiHoc*, 2007, pp. 32–40.

[12] K. Papanikolaou, C. X. Mavromoustakis, G. Mastorakis, A. Bourdena, and C. Dobre, "Energy consumption optimization using social interaction in the mobile cloud," *Mobile Netw. Manage.*, vol. 141, pp. 431–445, 2015.

[13] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. ACM SIGKDD*, 2001, pp. 57–66.

[14] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. ACM SIGKDD*, 2010, pp. 1029–1038.

[15] Y. Wang, A. V. Vasilakos, Q. Jin, and J. Ma, "PPRank: Economically selecting initial users for influence maximization in social networks," *IEEE Syst. J.*, pp. 1–12, 2015.

[16] A. Vahdat and D. Becker, "Epidemic routing for partially-connected ad hoc networks," Duke Univ., Durham, NC, USA, Tech. Rep. CS-200006, 2000.

[17] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Spray and wait: An efficient routing scheme for intermittently connected mobile networks," in *Proc. ACM SIGCOMM Workshop*, 2005, pp. 252–259.

[18] W. Zhao, Y. Chen, M. Ammar, M. D. Comer, B. N. Levine, and E. Zegura, "Capacity enhancement using throwboxes in DTNs," in *Proc. IEEE Int. Conf. Mobile Adhoc Sensor Syst. (MASS)*, 2006, pp. 31–40.

[19] F. Li and J. Wu, "Localcom: A community-based epidemic forwarding scheme in disruption-tolerant networks," in *Proc. 6th Annu. IEEE Commun. Soc. Conf. Sensor Mesh Ad Hoc Commun. Netw. (Secon)*, 2009, pp. 1–9.

[20] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 606–620, Jun. 2007.

[21] Y. Li, Y. Jiang, D. Jin, L. Su, L. Zeng, and D. Wu, "Energy-efficient optimal opportunistic forwarding for delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4500–4512, Nov. 2010.

[22] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Comput. Netw.*, vol. 51, no. 10, pp. 2867–2891, 2005.

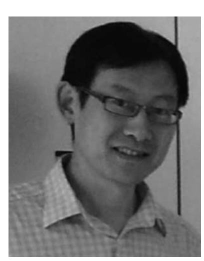
[23] C. Shao, S. Leng, Y. Zhang, A. Vinel, and M. Jonsson, "Performance analysis of connectivity probability and connectivity-aware MAC protocol design for platoon-based VANETs," *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5596–5609, Dec. 2015.

[24] Q. Wang, S. Leng, H. Fu, and Y. Zhang, "An IEEE 802.11p-based multichannel MAC scheme with channel coordination for vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 449–458, Jun. 2012.

[25] D. Kotz, T. Henderson, I. Abyzov, and J. Yeo. (2007). *CRAWDAD Data Set Dartmouth/Campus* (v. 2007-02-08) [Online]. Available: http://crawdad.org/dartmouth/campus/

Bo Fan (S'15) received the B.Eng. degree in communication engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2012, where he is currently pursuing the Ph.D. degree in communication and information engineering. His research interests include resource allocation and data delivery in mobile social networks.

Supeng Leng (M'06) received the Ph.D. degree in wireless networks from Nanyang Technological University (NTU), Singapore, in 2005. He is currently a Full Professor with the School of Communication and Information Engineering, UESTC. His research interests include broadband wireless access networks, vehicular networks, the next generation mobile networks, internet of things, mobile social networks, and smart grid.



Kun Yang (SM'08) received the Ph.D degree in wireless communications and networking from the University College London (UCL), London, U.K., in 2006. He is currently a Full Professor with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K. His research interests include heterogeneous wireless networks, fixed mobile convergence, pervasive service engineering, future Internet technology, network virtualization, and cloud computing.

Jianhua He (SM'12) received the Ph.D. degree in wireless communications and networking from Nanyang Technological University, Singapore, in 2002. He is currently a Lecturer with the School of Engineering and Applied Science, Aston University, Birmingham, U.K. His research interests include protocol design, modeling, and optimization for wireless network (cellular networks, wireless LAN, WiMax, wireless PAN, and vehicle networks), software defined network (SDN), smart cities, and Internet of Things.